

Improved Identification of Peptides by Enhanced Data Processing of High Resolution MALDI TOF/TOF Mass Spectra Prior to Database Searching

Tomas Rejtar, Hsuan-shen Chen, Viktor Andreev, Eugene Moskovets and Barry L. Karger

Barnett Institute and Department of Chemistry, Northeastern University, Boston, MA, USA

Tandem mass spectrometry followed by database (DB) searching has become well-established tools for peptide and the resultant protein identification. Typically, a list of fragment ions and their intensities, representing the tandem mass spectrum, is compared with a theoretically predicted fragmentation pattern for all peptides within the mass tolerance window of the precursor ion. We have developed a new approach for MS/MS spectra processing based on a sequential strategy, where the database search is performed on multiple peak lists created from a single MS/MS spectrum. This strategy overcomes the inherent disadvantages of using unified criteria for processing of all acquired spectra. Compared to the standard procedure, the new approach leads to an increase in the number of identified peptides, resulting in new protein identifications as well as more confident identification of the previously identified proteins.

A model sample of a tryptic digest of yeast soluble protein fraction separated by strong cation exchange (SCX) chromatography was used for evaluation of the enhanced procedure. The individual fractions were separated by reversed phase nanoLC and deposited on a standard MALDI plate in the form of continuous streaks. In a second round, a narrow streak of external standards was deposited in close proximity to the sample streak [1]. The deposited samples were analyzed by the AB 4700 MALDI TOF/TOF instrument in both the MS and MS/MS modes. The MS spectra were calibrated using standards deposited in close proximity to the sample trace, resulting in a mass accuracy of ± 10 ppm. The MS spectra were denoised by the MEND algorithm, which also selected precursors for ions for the MS/MS analysis [2]. After the MS/MS acquisition, all spectra were processed by software provided by Applied Biosystems as well as the in-house developed program using Mascot 1.9 database search engine.

Ten out of total 20 SCX fractions were analyzed resulting in 9,746 MS/MS spectra. All spectra were processed using standard instrument software and searched against a yeast DB using the Mascot algorithm. The results were considered as standard processing and used for comparison with the enhanced processing method. The enhanced method consisted of several steps, see Fig. 1. First, for each MS/MS spectrum, fragment ions within the mass forbidden ranges were removed, and three peak lists were generated using variable intensity thresholds. All three peak lists were then searched against the yeast DB with ± 10 ppm precursor ion mass tolerance, and all significant hits were stored in a relational database. Those spectra that did not yield significant Mascot score were subsequently searched with semitryptic enzyme specificity, and significant hits were again stored separately. The remaining MS/MS spectra were denoised using wavelet transformation, and the DB search was repeated again with additional significant hits included in the list of identified peptides. The enhanced processing using improved mass accuracy resulted in 37% more peptide identifications, see Table I, which represented about 22% new protein hits with at least 2 significant unique peptides, see Table II. In order to validate the enhanced procedure, the rate of false positive identifications for both types of processing was estimated by searching against the yeast DB with reversed protein sequences. The analysis showed that rate of false positive identifications was 2.7% and 3.8% for standard and enhanced processing, respectively. The increase in the false positive identifications was still below 5%, as referred by confidence interval defined by Mascot. Thus, it was concluded that the enhanced processing significantly decreased rate of false negative identifications and only slightly increased false positive rate. Even though the enhanced processing was applied to MALDI TOF/TOF MS spectra; generally, any high-resolution tandem mass spectra would be applicable.

References:

1. E. Moskovets, H.S. Chen, A. Pashkova, T. Rejtar, V.P. Andreev, and B.L. Karger. *Rapid Commun Mass Spectrom*, 2003 17(19) 2177-87.
2. V.P. Andreev, T. Rejtar, H.S. Chen, E.V. Moskovets, A.R. Ivanov, and B.L. Karger *Anal Chem*, 2003 75(22) 6314-26.

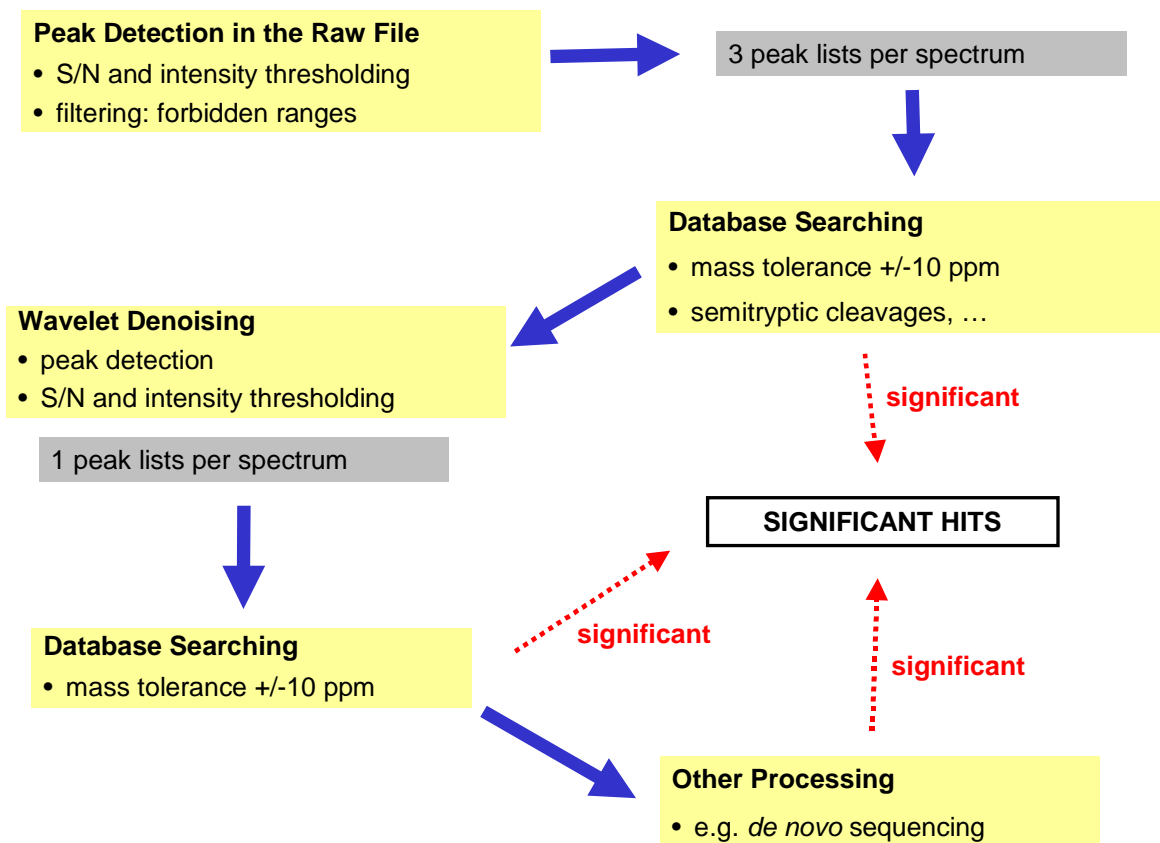


Figure 1. Scheme of the overall enhanced processing procedure

Table I. Comparison of the number of significant identifications in the normal and reversed database for standard and enhanced processing

Database	Standard	Enhanced Processing					
	Total ^a	Peak List process ^b	Mass Tolerance ^c	Semi ^d	Wavelet ^e	Total	New IDs
Normal	1777	201	176	96	178	2428	651(37%)
Reversed	48 (2.7%)	12	13	2	21	96 (3.9%)	48

a) number of identified peptides using trypsin specificity with +/-50 ppm precursor mass tolerance; b) additional identifications after generation of 3 peak lists per MS/MS spectrum with +/-50 ppm mass tolerance; c) additional identifications with +/-30 ppm and +/-10 ppm mass tolerance; d) additional significant peptides with only one tryptic end with +/-10 ppm mass tolerance; e) additional significant identifications after wavelet denoising using trypsin specificity with +/-10 ppm mass tolerance.

Table II. Comparison of number of identified peptide and proteins using standard and enhanced processing

Number	Standard Processing	Enhanced Processing	Added Identifications
Total peptide IDs	1,777	2,428	36%
Unique peptides	1,156	1,529	33%
Unique proteins	461	567	23%
Protein IDs (>1peptides)	221	271	22%